



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number:

0 625 572 A1

(12)

**EUROPEAN PATENT APPLICATION
published in accordance with Art.
158(3) EPC**

(21) Application number: 93921061.3

(51) Int. Cl.⁵: **C12N 15/10**

(22) Date of filing: **22.09.93**

(86) International application number:
PCT/JP93/01359

(87) International publication number:
WO 94/08001 (14.04.94 94/09)

(30) Priority: **25.09.92 JP 280932/92
03.03.93 JP 67589/93**

Applicant: **SEKINE, Shingo
4-4-1, Nishioonuma,
Sagamihara-shi
Kanagawa 229 (JP)**

(43) Date of publication of application:
23.11.94 Bulletin 94/47

(72) Inventor: **KATO, Seishi 1-9-2-304, Minamidai
Sagamihara-shi
Kanagawa 228 (JP)**
Inventor: **SEKINE, Shingo 4-4-1, Nishioonuma
Sagamihara-shi
Kanagawa 229 (JP)**

(71) Applicant: **THE KANAGAWA ACADEMY OF
SCIENCE AND TECHNOLOGY FOUNDATION
3-2-1, Sakado, Takatsuku,
Kawasaki-shi
Kanagawa 213 (JP)**
Applicant: **KATO, Seishi
1-9-2-304, Minamidai,
Sagamihara-shi
Kanagawa 228 (JP)**

(74) Representative: **Greaves, Carol Pauline et al
Mewburn Ellis
York House
23 Kingsway
London WC2B 6HP (GB)**

(54) **PROCESS FOR SYNTHESIZING COMPLETE-LENGTH cDNA, PROCESS FOR PRODUCING
INTERMEDIATE THEREFOR, AND PROCESS FOR PRODUCING RECOMBINANT VECTOR CONTAINING
COMPLETE-LENGTH cDNA.**

(57) The invention discloses a method of providing a cDNA containing the whole of the information on the primary structure of a protein by selectively synthesizing only the complete-length cDNA that contains a sequence beginning with the cap site of a mRNA. The process for producing an intermediate for the synthesis of the complete-length cDNA comprises the step of treating a mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped incomplete-length mRNA, the step of decapping from the 5' end of a capped complete-length mRNA,

and the step of linking the 5'-end phosphate group formed in the above step to either the DNA oligonucleotide represented by the following general formula (I) or a DNA-RNA chimera oligonucleotide by means of a T4RNA ligase to selectively add the DNA oligonucleotide or DNA-RNA chimera oligonucleotide having an arbitrary sequence to the 5' end of the complete-length mRNA: 5'-dN₁-dN₂-...-dN_m-N₁N₂-...-N_n-3', wherein dN represents a deoxyribonucleotide selected from among dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from among AMP, CMP, GMP and UMP; "-

" represents a phosphoric ester linkage; m represents an integer of 1 or above; and n represents an integer of 0 or above. The process for synthesizing a complete-length cDNA comprises linking a double-stranded DNA primer having a dT tail by annealing to the poly (A) tail of the 3' end of the complete-

length mRNA having the 5' end to which the DNA oligonucleotide or DNA-RNA chimera oligonucleotide have been added as described above, and then synthesizing a single-stranded cDNA complementary to the complete-length mRNA by means of a reverse transcriptase.

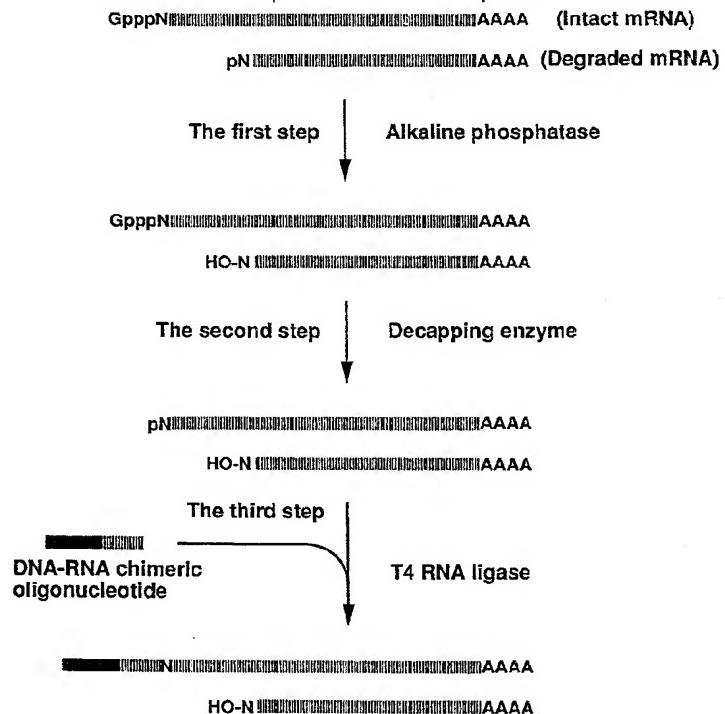


Fig.1

Technical Field

The present invention relates to a process for synthesizing a full-length cDNA from mRNA. This invention enables screening of proteins useful in industry and mass production of the proteins.

Background Art

Bio-active proteins produced by cells have been widely utilized in industry as materials for medicines, diagnosis, biosensors, bio-reactors and so on. The progress of gene technology has facilitated the discovery of these proteins and enabled us to produce a large amount of the proteins. The fundamental technology is a cDNA cloning technique.

Information on the amino acid sequence of a protein is encoded by mRNA. If the mRNA is converted to a corresponding DNA, i.e., a complementary DNA (cDNA), the cDNA can be used for determining the primary structure of the protein and producing a large amount of the protein. Therefore, there have been developed various cDNA cloning techniques which include isolating a mRNA from cells, synthesizing cDNAs from the mRNA, and isolating a cDNA encoding a target protein.

The most important requisite for cDNA cloning is to synthesize a cDNA containing an entire protein coding region encoded by an intact mRNA. The intact mRNA contains a characteristic structure at the 5' terminal, so called a cap. The cDNA containing an entire sequence involving from the nucleotide at which the cap is added is called a "full-length cDNA". If a full-length cDNA is synthesized, we can know the whole information on the primary structure of a protein and also use it for the mass production of the protein encoded by the cDNA. According to the conventional terminology, a "full-length cDNA" have been used as a cDNA containing an entire coding region of the protein even if the cDNA contains no cap site sequence. In this invention, however, a "full-length cDNA" is defined only as a cDNA containing a nucleotide sequence starting from the cap site.

The Gubler-Hoffman method [Gene 25:263-269(1983)], which is most widely used for cDNA synthesis, could not give a full-length cDNA, because deletion at the terminals of cDNA occurs. On the other hand, the Okayama Berg method [Mol.Cell.Biol. 2:161-170(1982)] is known to give a full-length cDNA at high efficiency. The characteristic of this method is to add a dC tail to the 3' end of the first strand cDNA synthesized from mRNA. This method, however, does not always give a full-length cDNA, because dC tailing also occurs to the truncated cDNA generated by unexpected termina-

tion of cDNA synthesis. The truncated cDNAs originated from the degraded mRNAs are also synthesized.

5 Disclosure of the Invention

It is therefore an object of the present invention to provide a method for selectively synthesizing a full-length cDNA containing a nucleotide sequence starting from a cap site of mRNA, and a cDNA containing whole information on the primary structure of a protein. Furthermore another object of this invention is to provide a method to prepare an intermediate for synthesizing the full-length cDNA and to prepare a recombinant vector containing the full-length cDNA.

10 The inventors succeeded in synthesizing a full-length cDNA at high efficiency from mRNA whose cap structure was replaced by a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide.

20 The invention provides a process for producing an intermediate for the synthesis of a full-length cDNA, which comprises a step of treating mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped degraded mRNA, a step of decapping from the 5' end of a capped intact mRNA, and a step of ligating either a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide represented by the following general formula [I] to the 5' end phosphate group formed in the above step by the action of T4 RNA ligase, thereby selectively adding either the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide having an arbitrary sequence to the 5' end of the intact mRNA.

25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 9999

cDNA described above in which the sequence of the DNA oligonucleotide or the DNA-RNA oligonucleotide contains at least one site recognized by a restriction enzyme RE1 and the double-stranded DNA primer contains the same, the step of digesting with a restriction enzyme RE1 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by above method, the step of circularizing the resultant material by self-ligation, and the step of converting the RNA in the circularized recombinant vector to DNA.

The invention enables sure synthesis of a full-length cDNA which has been difficult to be synthesized according to the conventional methods. Since a cDNA clone obtained from the cDNA library prepared according to the present method is sure to contain the whole information on the primary structure of the protein, the obtained clone can be used immediately to produce the encoded protein.

Brief Description of the Drawings

Figure 1 shows a process to selectively ligate a DNA-RNA chimeric oligonucleotide to an intact mRNA possessing a cap; and Figure 2 shows a process to synthesize cDNA from a synthetic oligonucleotide-ligated mRNA.

Best Mode for Carrying Out the Invention

This invention is concerned with a method for synthesizing a full-length CDNA, including two processes. The first process consists of ligation of a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide to an intact mRNA possessing a cap and the second process consists of cDNA synthesis from the synthetic oligonucleotide-capped mRNA and production of a recombinant vector containing the cDNA. Each step in the first process is shown in Figure 1. All steps in the first and second processes can be done using the conventional methods in this field.

The First Step

Usually mRNAs extracted from cells contain degraded ones possessing no cap structure. This step is carried out to remove the phosphate group of the degraded mRNA by means of an enzyme such as an alkaline phosphatase. Since an intact mRNA has a cap structure at the 5' end, the phosphatase does not affect the intact mRNA. The 5' end of a degraded mRNA possessing no cap structure is converted to a hydroxy group by selective dephosphorylation.

The Second Step

5 The cap structure of the intact mRNA is removed by a decapping enzyme to generate one phosphate group at the 5' end. The tobacco acid pyrophosphatase or T4 polynucleotide kinase can be exemplified as the decapping enzyme.

The Third Step

10 The DNA oligonucleotide or DNA-RNA chimeric oligonucleotide described by the following formula [I] is ligated to the 5'-terminal phosphate group of the decapped intact mRNA.

15
$$5'-dN_1-dN_2-\cdots-dN_m-N_1-N_2-\cdots-N_n-3'$$
 [I]
wherein dN represents a deoxyribonucleotide selected from among dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from among AMP, CMP, GMP, and UMP; "-" represents a phosphoric ester linkage; m represents an integer of 1 or above; n represents an integer of 0 or above.

20 In order to facilitate the followed preparation of a recombinant vector containing the cDNA, the sequence of the deoxyribonucleotide preferably contains at least one site to be recognized by a restriction enzyme RE1. Any restriction enzyme RE1 which does not cut the DNA-RNA hybrid can be used, and one generating the protruding site is preferred. In this step the 5' end of the intact mRNA, which has a phosphate group generated at the above second step, can incorporate with a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide. On the other hand, the 5' end of the degraded mRNA cannot incorporate with it because of the hydroxyl group generated at the 5' end by dephosphorylation in the first step. Thus the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide is selectively added to the 5' end of the intact mRNA. Since T4 RNA ligase catalyzes the ligation between RNA strands more efficiently than between RNA and DNA strands, the use of a DNA-RNA chimeric oligonucleotide is preferable.

25 30 35 40 45 The DNA oligonucleotide or DNA-RNA chimeric oligonucleotide used here can be chemically synthesized in a tube or with an automatic DNA/RNA synthesizer. There have been no reports on the DNA-RNA chimeric oligonucleotide in spite of its easy synthesis, so that it is provided by this invention for the first time.

45 It is preferable to use excess molar ratio of the oligonucleotide to mRNA in the ligation reaction between the oligonucleotide and mRNA.

50 55 Each step of the second process is shown in Figure 2.

The First Step

The mRNA capped by a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide, which was prepared in the first process, is annealed with a double-stranded DNA primer containing a tail of n (dT)s at the 3' end. Then the first strand cDNA is synthesized using reverse transcriptase. The double-stranded DNA primer can be selected from an oligonucleotide, any double-stranded DNA fragment or any vector primer (for example, a plasmid DNA containing a replication origin of E.coli). The oligonucleotide primer can be chemically synthesized. In this case, n is preferably from 10 to 30. The latter two primers can be prepared by cutting the plasmid vector with a restriction enzyme generating a protruding 3' end and adding a dT-tail using terminal deoxynucleotidyl transferase. In this case, n is preferably from 50 to 70.

The double-stranded DNA primer preferably possesses at least one site for restriction enzyme RE2 near the end to which the dT-tail is not added for easy preparation of a recombinant vector containing a cDNA. Any restriction enzyme RE2, preferably one generating a protruding end by digestion, can be used as long as it does not cut a mRNA-cDNA hybrid. The RE1 and RE2 should be identical when a vector primer is used, although they could be different when an oligonucleotide primer or a DNA fragment primer is used.

The synthesis of the first strand cDNA complementary to a intact mRNA is completed by the first step described above. The following second, third, and fourth steps are those for preparing a recombinant vector containing the synthesized cDNA.

The Second Step

When two restriction sites for RE1 and RE2 are used, the synthesized product containing the first strand cDNA is digested with both RE1 and RE2 to generate the corresponding restriction end at both ends of the double-stranded DNA parts. When no restriction sites exist in these parts, any restriction end can be prepared by adding linkers to the ends with T4 DNA ligase and digesting them with the corresponding restriction enzymes according to the conventional method. The blunt ends can be used without digestion with restriction enzymes.

The Third Step

This is a step for inserting the obtained cDNA into an appropriate cloning vector for E.coli. The procedure depends on the used primer. When an oligonucleotide primer or a DNA fragment primer is used, the cDNA fragment prepared in the second

step is ligated to any cloning vector (including plasmid, cosmid, phage and so on) for E.coli after digesting it with restriction enzymes RE1 and RE2. If a linker is used for preparing the restriction ends in the second step, the vector should be digested with the restriction enzymes corresponding to the site designed in the linker, or the corresponding linker should be added to the ends of the vector prior to the ligation to the cDNA fragment. In either cases, the ligation between blunt ends can be carried out without digestion with restriction enzymes.

The Fourth Step

The RNA strand in the vector is replaced by a DNA strand using RNase H, E.coli DNA polymerase I, and E.coli DNA ligase according to the conventional method.

Above procedure allows to prepare a recombinant vector containing a cDNA complementary to an intact mRNA.

The recombinant vector prepared by the above procedure can be used to transform host cells such as E.coli. The plasmid, cosmid, and phage etc. prepared from the transformants contain the full-length cDNA. Thus these processes give a clone containing a full-length cDNA.

The present invention will now be described by way of examples, but various changes and modifications can be made without departing from the invention.

REFERENCE EXAMPLES

(1) Preparation of mRNA

A model mRNA was prepared from the cDNA encoding human ribosomal phosphoprotein P1 (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292). Five µg of the cDNA plasmid was digested with NotI and then used as a template for in vitro transcription using a kit purchased from Boehringer Mannheim. By adding m7G(5')ppp(5')G in the reaction mixture, the cap was added to the 5' end of the transcribed RNA. When the product was analyzed by formamide-containing agarose gel electrophoresis, a single band of c.a. 600 nucleotides was shown on the gel. This was used as a model mRNA for cDNA synthesis. This model mRNA is composed of the vector-derived sequence starting from the cap, the sequence derived from the mRNA encoding human ribosomal phosphoprotein P1, and about 100 nucleotides of poly(A) tail. Not all mRNAs synthesized above contain the cap at the 5' end.

(2) Synthesis of DNA-RNA chimeric oligonucleotide

The oligonucleotide described by the following formula was synthesized with a DNA synthesizer (Applied Biosystems, model 392) using the phosphoramidite method.

5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-G-G-A-3'

Reagents for RNA synthesis and DNA synthesis were purchased from Peninsula and ABI, respectively. First the adenine-CPG column was set on the synthesizer. Then two ribonucleotides (G) followed by the other deoxyribonucleotides were added in order of above sequence under the reaction conditions described in the protocol. Finally the synthesized oligomer was cut in an ammonia-ethanol (3:1) solution. The protected groups were removed by treating with THF at 55°C for 10 h. The product was purified on polyacrylamide gel electrophoresis.

(3) Synthesis of DNA oligonucleotide

The DNA oligonucleotide described by the following formula was synthesized with a DNA synthesizer (Applied Biosystems, model 392) using the phosphoramidite method.

5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-dG-dG-dA-3'

The reaction conditions described in the protocol were used. The product was purified on polyacrylamide gel electrophoresis according to the conventional method.

(4) Preparation of the DNA fragment primer

One hundred µg of the pKA1 vector primer possessing a tail of c.a. 60 (dT)s (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292) was digested with 100 units of the restriction enzyme NdeI. The 500 bp fragment containing a (dT) tail was isolated on 1.5 % agarose gel electrophoresis. The resultant fragment was used as a DNA fragment primer.

EXAMPLE 1

One hundred µg of the model mRNA encoding human ribosomal phosphoprotein P1, which was prepared in above REFERENTIAL EXAMPLE, was dissolved in 100 mM Tris-HCl (pH 8.0). After adding one unit of RNase-free bacterial alkaline phosphatase (Takara Shuzo), the reaction mixture was incubated at 37°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in a solution containing 50 mM sodium acetate (pH 6.0), 1 mM EDTA, 0.1 % mercaptoethanol, and 0.01% Triton

X-100. After adding one unit of tobacco acid phosphatase (Epicentre Technologies), the reaction mixture of total 100 µl was incubated at 37°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

5 Fifteen pmoles of the decapped mRNA and 150 pmoles of the DNA-RNA chimeric oligonucleotide synthesized in above REFERENTIAL EXAMPLE were dissolved in a 100 µl solution containing 75 mM Tris-HCl (pH 7.5), 0.1 mM ATP, 10 mM MgCl₂, 5 mM dithiothreitol, and 10 µl of DMSO. After adding 100 units of T4 RNA ligase (Takara Shuzo), the reaction mixture was incubated at 16°C for 16 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

10 Fifteen µg of the DNA-RNA chimeric oligonucleotide-ligated mRNA and 1.5 µg of a pKA1 vector primer possessing a tail of c.a. 60 (dT)s (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292) were dissolved in a solution containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, 1.25 mM dNTPs(dATP + dCTP + dGTP + dTTP). After adding 200 units of reverse transcriptase (Takara Shuzo), the reaction mixture of total 20 µl was incubated at 42°C for one hour. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in a solution containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 4 mM MgCl₂, 10 mM (NH₄)₂SO₄, and 50 µg/ml bovine serum albumin. Then the reaction mixture was added by 60 units of E.coli DNA ligase (Takara Shuzo) and incubated at 16°C for 16 hours. After adding 2 µl of 2 mM dNTPs, 4 units of E.coli DNA polymerase (Takara Shuzo) and 0.1 unit of E.coli RNase H (Takara Shuzo), the reaction mixture was incubated at 12°C for one hour and then at 22°C for one hour.

15 Mixing 10 µl of the reaction mixture with 100 µl of E.coli HB101 competent cells, the transformation of HB101 was carried out according to the conventional method. The transformants grew on an ampicillin-containing LB agar plate. Twelve colonies generated on the plate were picked up, suspended in 3 ml of an ampicillin-containing LB medium, and incubated overnight. The plasmids were isolated from the cultured cells by the alkaline lysis method. The isolated plasmids were digested with EcoRI and NotI and analyzed on agarose gel electrophoresis. Each plasmid contained a cDNA insert of c.a. 600 bp. The sequence of the cDNA terminal was determined by the dideoxy method. The results showed that each cDNA possessed the sequence starting from the cap site of the model mRNA downstream of the synthetic oligonucleotide-derived sequence including an Ec-

oRI site.

EXAMPLE 2

Five µg of rabbit globin mRNA (BRL) was dephosphorylated and decapped by the same way as described in EXAMPLE 1. Thirty pmoles of the obtained mRNA and 3 nmoles of the DNA-RNA chimeric oligonucleotide synthesized in REFERENTIAL EXAMPLE were dissolved in a solution containing 50 mM Tris-HCl (pH 7.5), 0.5 mM ATP, 5 mM MgCl₂, 10 mM 2-mercaptoethanol, and 25 % polyethylene glycol. After adding 50 units of T4 RNA ligase (Takara Shuzo), the reaction mixture of total 30 µl was incubated at 20 °C for 12 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

After the reaction product was annealed with a pKA1 vector primer, cDNA synthesis and transformation of E.coli by the resultant cDNA vector were carried out under the same condition as described in EXAMPLE 1. Twelve colonies were randomly picked up from the agar plate. After incubation in the medium, the plasmid was isolated from the cultured cells. Since a rabbit globin cDNA has one EcoRI site at position c.a. 400 from the cap site, the full-length cDNA encoding rabbit globin should produce the fragment of 400 bp by EcoRI digestion. In fact, the double digestion of every plasmid with EcoRI and NotI produced two bands including c.a. 400 bp and c.a. 250 bp originated from the cDNA and one band corresponding to the vector. Those results suggested that every plasmid contained the full-length cDNA for globin. In order to confirm this, the 5' terminal sequences were determined. The results showed that each cDNA possessed the sequence starting from the cap site of the rabbit globin mRNA downstream of the synthetic oligonucleotide-derived sequence including an EcoRI site. The starting material, rabbit globin mRNA, is composed of alpha-globin mRNA and beta-globin mRNA. Out of 12 clones analyzed, 4 clones contained the full-length cDNA encoding alpha-globin and the remaining 8 clones contained that encoding beta-globin.

EXAMPLE 3

Using a DNA oligonucleotide prepared in REFERENCCE EXAMPLE 3 instead of a DNA-RNA chimeric oligonucleotide, cDNA synthesis was done in the same way as described in above example. The number of transformants decreased by one tenth, but the sequence analysis of 5' end of cDNA isolated from the transformants showed that every clone contained the full-length cDNA encoding rabbit alpha- or beta-globin.

EXAMPLE 4

Three µg of the DNA-RNA chimeric oligonucleotide-ligated rabbit globin mRNA prepared in EXAMPLE 2 was annealed with 1.5 µg of the DNA fragment primer prepared in REFERENCE EXAMPLE, and then cDNA synthesis was carried out under the same condition as described in EXAMPLE 1. The product was digested with 50 units of EcoRI and 50 units of NotI, and ligated to 1 µg of EcoRI,NotI-digested pKA1 (described in Japanese Laid-open Patent Application (Kokai) No. 4-117292). The reaction mixture was used for transformation of E.coli. The plasmid was prepared from each transformant. The double digestion of every plasmid with EcoRI and NotI produced two bands including c.a. 400 bp and c.a. 250 bp on agarose gel electrophoresis. The 5' terminal sequences were determined by the dideoxy method. The results showed that each clone possessed the full-length cDNAs encoding the rabbit alpha- or beta-globin downstream of the synthetic oligonucleotide-derived sequence including an EcoRI site.

EXAMPLE 5

Ten µg of poly(A)⁺RNA prepared from human fibrosarcoma cell line HT-1080 using the conventional method was dephosphorylated and then decapped in the same way as described in EXAMPLE 1. The obtained poly(A)⁺RNA and 3 nmoles of the DNA-RNA chimeric oligonucleotide synthesized in REFERENCE EXAMPLE were dissolved in a solution containing 50 mM Tris-HCl (pH 7.5), 0.5 mM ATP, 5 mM MgCl₂, 10 mM 2-mercaptoethanol, and 25 % polyethylene glycol. After adding 50 units of T4 RNA ligase (Takara Shuzo), the reaction mixture of total 30 µl was incubated at 20 °C for 12 hours. After phenol extract and ethanol precipitation of the reaction mixture, the pellet was dissolved in distilled water.

After the reaction product was annealed with the PKA1 vector primer, cDNA synthesis and transformation of E.coli by the resultant cDNA vector were carried out under the same condition as described in EXAMPLE 1. Six hundred colonies were randomly picked up from the agar plate. After incubation in the medium, the plasmids were isolated from the cultured cells. Determining the 5' terminal sequence of each cDNA clone, we found out that many cDNAs encoded proteins identical to known human proteins. In some genes encoding these proteins, their transcription initiation sites on the genomic sequence have already been determined, so that we compared it to the 5' terminal sequence of the obtained cDNA. In the case of abundant housekeeping proteins such as elongation factor 1-alpha, various ribosomal proteins and

so on, the 5' terminal sequence of the cDNA agreed with the sequence of the transcription initiation site. These results mean that the cDNA is full-length. Nine out of 10 clones encoding elongation factor 1-alpha contained the full-length cDNA (starting from 5'-CTTTTCGCAA...).

INDUSTRIAL AVAILABILITY

The invention enables sure synthesis of a full-length cDNA. Since a cDNA clone obtained from the cDNA library prepared according to the present method is sure to contain the whole information on the primary structure of the protein, the obtained clone can be used immediately to produce the encoded protein. Therefore, the present invention is useful for production of useful proteins by genetic engineering.

Claims

1. A process for producing an intermediate for the synthesis of a full-length cDNA, which comprises the steps of treating mRNA extracted from cells with an alkaline phosphatase to eliminate the phosphate group from the 5' end of an uncapped degraded mRNA; decapping from the 5' end of a capped intact mRNA; and ligating either a DNA oligonucleotide or a DNA-RNA chimeric oligonucleotide represented by the following general formula [I] to the 5' end phosphate group formed in the above step by the action of T4 RNA ligase, thereby selectively adding either the DNA oligonucleotide or the DNA-RNA chimeric oligonucleotide having an arbitrary sequence to the 5' end of the intact mRNA.

5'-dN1-dN2-···-dNm-N1-N2-···-Nn-3' [I]

(wherein dN represents a deoxyribonucleotide selected from dAMP, dCMP, dGMP and dTMP; N represents a ribonucleotide selected from AMP, CMP, GMP and UMP; "-" represents a phosphoric ester linkage; m represents an integer of 1 or more; and n represents an integer of 0 or more.)

2. The process according to claim 1, wherein said DNA oligonucleotide or said DNA-RNA chimeric oligonucleotide represented by the formula [I] contains at least one recognition site of a restriction enzyme RE1 which does not cleave mRNA-cDNA hybrid which is formed in a subsequent step.
3. A process for synthesizing a full-length cDNA, which comprises linking a double-stranded DNA primer having a dT tail by annealing to a poly(A) tail of the 3' end of the intact mRNA

5 having the DNA oligonucleotide or the DNA-RNA chimeric nucleotide added at the 5' end produced by the process according to claim 1, and then synthesizing the first strand cDNA complementary to the intact mRNA by a reverse transcriptase.

4. A process for synthesizing a full-length cDNA, which comprises linking a double-stranded DNA primer having a dT tail by annealing to a poly(A) tail of the 3' end of the intact mRNA having the DNA oligonucleotide or the DNA-RNA chimeric nucleotide added at the 5' end produced by the process according to claim 2, and then synthesizing the first strand cDNA complementary to the intact mRNA by a reverse transcriptase.
5. A process for preparing a recombinant vector containing a full-length cDNA, which comprises the steps of digesting with the restriction enzyme RE1 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by the process according to claim 4 wherein said double-stranded DNA primer is a vector primer containing at least one recognition site of said restriction enzyme RE1; circularizing the resultant material by self-ligation; and converting the RNA in the circularized recombinant vector to DNA.
6. A process for preparing a recombinant vector containing a full-length cDNA, which comprises the steps of digesting with the restriction enzyme RE1 and a restriction enzyme RE2 the material linked between the double-stranded DNA primer and the mRNA-cDNA hybrid prepared by the process according to claim 4 wherein said double-stranded DNA primer is a vector primer containing at least one recognition site of said restriction enzyme RE2; recombining the resultant material with a vector having recognition sites of restriction enzymes RE1 and RE2 at its ends, respectively, by ligation; and converting the RNA in the circularized recombinant vector to DNA.
7. A DNA-RNA chimeric oligonucleotide.

50

55



Fig.1

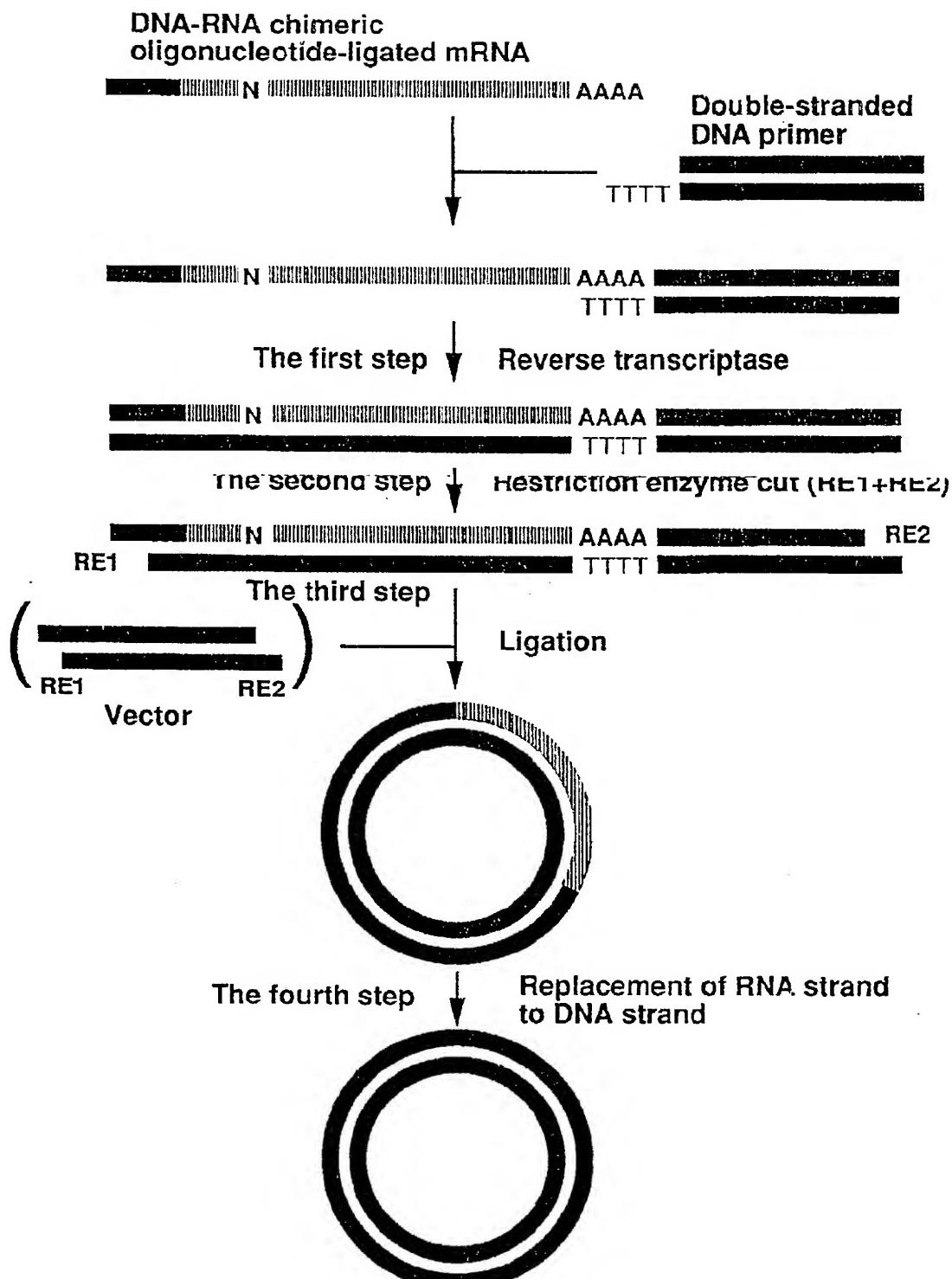


Fig.2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP93/01359

A. CLASSIFICATION OF SUBJECT MATTER

Int. C15 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. C15 C12N15/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS PREVIEWS, CAS ONLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Gene, Vol. 68, 1988, Rutledge RG et. al. "Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors" p. 151-158	1-7

 Further documents are listed in the continuation of Box C. See patent family annex.

- * Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
December 21, 1993 (21. 12. 93)

Date of mailing of the international search report

February 8, 1994 (08. 02. 94)

Name and mailing address of the ISA/
Japanese Patent Office
Facsimile No.

Authorized officer

Telephone No.